

April 3, 2014

# TWO-SAMPLE U-STATISTIC PROCESSES FOR LONG-RANGE DEPENDENT DATA

HEROLD DEHLING, AENEAS ROOCH, AND MARTIN WENDLER

**ABSTRACT.** Motivated by some common change-point tests, we investigate the asymptotic distribution of the U-statistic process

$$U_n(t) = \sum_{i=1}^{[nt]} \sum_{j=[nt]+1}^n h(X_i, X_j), \quad 0 \leq t \leq 1,$$

when the underlying data are long-range dependent. We present two approaches, one based on an expansion of the kernel  $h(x, y)$  into Hermite polynomials, the other based on an empirical process representation of the U-statistic. Together, the two approaches cover a wide range of kernels, including all kernels commonly used in applications.

## 1. INTRODUCTION

In this paper, we will investigate the asymptotic distribution of the two-sample U-statistic process, defined as

$$(1) \quad U_n(t) = \sum_{i=1}^{[nt]} \sum_{j=[nt]+1}^n h(X_i, X_j), \quad 0 \leq t \leq 1,$$

when the underlying data  $(X_i)_{i \geq 1}$  are long-range dependent. We will present two general approaches to the analysis of the two-sample U-statistic process, one based on the Hermite expansion of the kernel  $h(x, y)$ , and the other on an empirical process representation. Combined, the two approaches cover a wide range of kernels, including all kernels commonly applied in change-point tests.

Two-sample U-statistic processes find applications in the detection of change-points in a time series  $(X_i)_{i \geq 1}$ . Many common test statistics for changes in the location can be expressed in the form  $\max_{1 \leq k \leq n-1} \sum_{i=1}^k \sum_{j=k+1}^n h(X_i, X_j)$ . This holds, for example, for the CUSUM test and the Wilcoxon change-point test, where  $h(x, y) = y - x$  and  $h(x, y) = 1_{\{x \leq y\}}$ , respectively. The asymptotic distribution of such test statistics can be obtained if one knows the limit distribution of the corresponding two-sample U-statistic process.

The asymptotic distribution of the two-sample U-statistic process has been obtained earlier by Csörgő and Horváth (1988), for i.i.d. data, and by Dehling, Fried, Garcia and Wendler (2013) for short-range dependent data. In both cases,  $n^{-3/2}(U_n(t))_{0 \leq t \leq 1}$  converges in distribution, on the space  $D[0, 1]$ , towards a Gaussian process. In the case of long-range dependent data, the two-sample U-statistic process has been studied only for two specific examples,

---

*Key words and phrases.* U-statistic processes, Long-range dependence, Hermite polynomials, Empirical processes, Change-point problems, Non-central limit theorems.

Research supported by Collaborative Research Project SFB 823 *Statistical Modeling of Nonlinear Dynamic Processes* of the German Research Foundation DFG.

namely for  $h(x, y) = y - x$  and  $h(x, y) = 1_{\{x \leq y\}}$ ; see Horváth and Kokoszka (1997) and Dehling, Rooch and Taqqu (2013a), respectively.

In this work, we consider a stationary Gaussian processes  $(\xi_i)_{i \geq 1}$  with mean zero, variance 1 and auto-covariance function

$$(2) \quad \gamma(k) = \text{Cov}[\xi_1, \xi_{1+k}] = L(k)k^{-D},$$

with  $0 < D < 1$  and a slowly varying function  $L$ . Note that  $\gamma_k$  obeys a power law, while short-range dependent processes typically possess an auto-correlation function that decays exponentially fast. In long-range dependent time series, i.e. time series with such slowly decaying auto-correlations, even observations in the distant past affect present behaviour; this is why long-range dependence is also called long memory.

Long-range dependence/long memory is an important issue in many fields: it can be detected in hydrology (Mandelbrot and Wallis, 1968) and in climate science (Caballero et al., 2002). Moreover, it is an omnipresent property of data traffic both in local area networks and in wide area networks, and it can be explained by renewal processes that exhibit heavy-tailed interarrival distributions (Levy and Taqqu, 2000) or by the superposition of many highly variable sources. For a survey see Willinger et al. (1996) and Cappé et al. (2002). In economics and finance, it is controversially discussed whether there is long-range dependence in economic data (Lo, 1991); volatilities can be long-range dependent processes (Breidt et al., 1998), and sometimes there is evidence of long-range dependence in some stock market prices, while sometimes there is none (Willinger et al., 1999). Baillie (1996) provides a survey of the major econometric work on long-range dependence, fractional integration and their application in economics. A short overview about probabilistic foundations and statistical models for long-range dependent data including extensive references is given by Beran (2010).

## 2. DEFINITIONS AND MAIN RESULTS

**2.1. Two-sample  $U$ -statistics processes.** Given two samples  $X_1, \dots, X_m$  and  $Y_1, \dots, Y_n$ , and a kernel  $h(x, y)$ , the (non-normalized) two-sample  $U$ -statistic is defined as

$$U_{m,n} = \sum_{i=1}^m \sum_{j=1}^n h(X_i, Y_j).$$

A well-known example is the Wilcoxon-Mann-Whitney test statistic for a difference in location between the two samples, which is obtained by taking  $h(x, y) = 1_{\{x \leq y\}}$ . Another example is the two-sample Gauss test statistic, which corresponds to  $h(x, y) = y - x$ .

In the present paper, we start from a single sample  $\xi_1, \dots, \xi_n$ , which is split into two pieces of consecutive observations  $\xi_1, \dots, \xi_{[n\lambda]}$ , and  $\xi_{[n\lambda]+1}, \dots, \xi_n$ , for any  $\lambda \in [0, 1]$ .

**Definition 2.1.** Let  $(\xi_i)_{i \geq 1}$  be a stochastic process, and let  $h : \mathbb{R}^2 \rightarrow \mathbb{R}$  be a measurable kernel. We define the two-sample  $U$ -statistic process

$$(3) \quad U_n(\lambda) = \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^n h(\xi_i, \xi_j), \quad 0 \leq \lambda \leq 1.$$

We will view  $(U_n(\lambda))_{0 \leq \lambda \leq 1}$  as an element of the function space  $D[0, 1]$ .

In what follows, we shall first derive the asymptotic distribution of this process, in the case when  $(\xi_i)_{i \geq 1}$  is a stationary Gaussian process with mean zero, variance 1, and autocovariance function (2). The general case of a Gaussian subordinated process, i.e.  $X_i = G(\xi_i)$ , follows by considering the transformed kernel  $h(G(x), G(y))$ .

For our results, we will usually require the kernels to satisfy moment conditions with respect to the standard normal distribution  $\mathcal{N} = \frac{1}{2\pi} e^{-x^2/2} e^{-y^2/2} dx dy$ . We denote by  $L^p(\mathbb{R}^2, \mathcal{N})$  the space of all  $p$ -integrable kernels. Observe that we may assume without loss of generality that  $h$  is centered, i.e. that  $E[h(\xi, \eta)] = 0$  for two i.i.d. Gaussian random variables  $\xi, \eta \sim \mathcal{N}(0, 1)$  – otherwise just subtract the mean.

One-sample U-statistics of LRD data have been studied by Dehling and Taqqu (1989, 1991) and by Lévy-Leduc et al. (2011). The two-sample U-statistic process has only been studied for some specific kernels. For the kernel  $h(x, y) = y - x$ , the asymptotic distribution can be obtained directly from the functional non-central limit theorem of Dobrushin and Major (1979) and Taqqu (1979); see Dehling, Roach and Taqqu (2013a) for details. For the kernel  $h(x, y) = 1_{\{x \leq y\}}$ , Dehling, Roach and Taqqu (2013a) obtained the asymptotic distribution of the two-sample U-statistic process using the empirical process non-CLT of Dehling and Taqqu (1989). As an application, they derived the asymptotic distribution of the Wilcoxon change-point test statistic for processes with LRD noise.

In this paper, we will derive the limit distribution of  $(U_n(\lambda))_{0 \leq \lambda \leq 1}$  for a broad class of kernels  $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ . We shall use two different approaches. The first approach uses a bivariate Hermite expansion of the kernel  $h(x, y)$ , while the second approach is based on an empirical process representation of the two-sample U-statistic process. In both cases, different technical assumptions regarding the kernel  $h(x, y)$  have to be made. Together, the two approaches cover most examples that are relevant in statistics.

**2.2. A direct approach via the Hermite expansion.** The motivation for the direct approach via the Hermite expansion arises from the study of partial sums of Gaussian subordinated processes. For any integer  $k \geq 0$ , we introduce the  $k$ -th order Hermite polynomial

$$(4) \quad H_k(x) = (-1)^k e^{x^2/2} \frac{d^k}{dx^k} e^{-x^2/2}.$$

It is well-known that the Hermite polynomials are orthogonal in the space  $L^2(\mathbb{R}, N)$ , and that the normalized Hermite polynomials  $H_k(x)/\sqrt{k!}$  form an orthonormal basis for  $L^2(\mathbb{R}, N)$ . Thus, any  $L^2$ -function  $h(x)$  can be expanded into a Hermite series

$$(5) \quad h(x) = \sum_{k=0}^{\infty} \frac{a_k}{k!} H_k(x),$$

where the coefficients are given by  $a_k = E(h(\xi) H_k(\xi))$ . Convergence in (5) is in the  $L^2$ -sense. We define the Hermite rank  $m$  of  $h$  as the index of the lowest order non-vanishing term in the Hermite expansion, i.e.

$$(6) \quad m = \min\{k \geq 0 : a_k \neq 0\}.$$

In this way, we may rewrite the Hermite expansion of  $h$  as

$$(7) \quad h(x) = \sum_{k=m}^{\infty} \frac{a_k}{k!} H_k(x).$$

Thus, we obtain the following series expansion for the partial sum of the  $h(\xi_i)$ ,

$$\sum_{i=1}^n h(\xi_i) = \sum_{k=m}^{\infty} \frac{a_k}{k!} \sum_{i=1}^n H_k(\xi_i).$$

The variance of the different terms in this expansion depends crucially on the index  $k$ . In fact, one obtains

$$(8) \quad \text{Var} \left[ \sum_{i=1}^n H_k(\xi_i) \right] \sim \begin{cases} nk!C, & \text{if } Dk > 1, \\ n^{2-Dk} L(n)^k \frac{2k!}{(1-Dk)(2-Dk)}, & \text{if } Dk < 1. \end{cases}$$

Moreover, for different indices  $k, l$ , the random variables  $H_k(\xi_i)$  and  $H_l(\xi_j)$  are uncorrelated, and thus

$$\text{Var} \left( \sum_{k=m}^{\infty} \frac{a_k}{k!} \sum_{i=1}^n H_k(\xi_i) \right) = \sum_{k=m}^{\infty} \frac{a_k^2}{(k!)^2} \text{Var} \left( \sum_{i=1}^n H_k(\xi_i) \right).$$

When  $mD < 1$ , this expansion is dominated by the lowest order term. In fact, defining

$$(9) \quad d_n^2 = d_n(m)^2 = \text{Var} \left( \sum_{i=1}^n H_m(\xi_i) \right) \sim c_m n^{2-Dm} L^m(n),$$

where

$$(10) \quad c_m = \frac{2m!}{(1-Dm)!(2-Dm)!},$$

Taqqu (1975) proved the reduction principle, stating that

$$\frac{1}{d_n} \left| \sum_{i=1}^n h(\xi_i) - \frac{a_m}{m!} \sum_{i=1}^n H_m(\xi_i) \right| \rightarrow 0,$$

in probability. Thus the study of the partial sums of arbitrary functionals of an LRD Gaussian process can be reduced to the study of partial sums of Hermite polynomials  $\sum_{i=1}^n H_m(\xi_i)$ . Weak convergence of the latter sums has been studied by Taqqu (1979), and independently by Dobrushin and Major (1979). These authors proved

$$(11) \quad \frac{1}{d_n} \sum_{i=1}^n H_m(\xi_i) \rightarrow Z_m(t),$$

where  $Z_m(t)$  denotes the  $m$ -th order Hermite process. For details, see e.g. the forthcoming monograph by Pipiras and Taqqu (2014).

Motivated by the results of Taqqu (1977, 1979) for partial sums of Gaussian subordinated processes, we will now study the Hermite expansion of functions  $h \in L^2(\mathbb{R}^2, \mathcal{N})$ . We define the two-dimensional Hermite polynomials  $H_{kl}(x, y) = H_k(x)H_l(y)$ , where  $H_k(x)$  is the one-dimensional Hermite polynomial, as defined in (4). Then,

$$\frac{H_k(x)H_l(y)}{\sqrt{k!l!}}, \quad k \geq 0, l \geq 0,$$

is an orthonormal basis for the Hilbert space  $L^2(\mathbb{R}^2, \mathcal{N})$ . Thus, we obtain for any  $h \in L^2(\mathbb{R}^2, \mathcal{N})$  the Hermite expansion

$$(12) \quad h(x, y) = \sum_{k, l=0}^{\infty} \frac{a_{kl}}{k! l!} H_k(x) H_l(y),$$

where the coefficients are given by

$$(13) \quad a_{kl} = E[h(\xi, \eta) H_k(\xi) H_l(\eta)] = \int_{\mathbb{R}^2} h(x, y) H_k(x) H_l(y) \varphi(x) \varphi(y) dx dy.$$

Here,  $\varphi(x)$  denotes the one-dimensional standard normal probability density function. Note that (12) is an expansion in the Hilbert space  $L^2(\mathbb{R}^2, \mathcal{N})$ , i.e. the series (12) converges in  $L^2(\mathbb{R}^2, \mathcal{N})$  towards the function  $h(x, y)$ .

We now order the terms in the expansion (12) according to the size of  $k + l$ :

$$(14) \quad h(x, y) = \sum_{q=m}^{\infty} \sum_{\substack{k, l: \\ k+l=q}} \frac{a_{kl}}{k! l!} H_k(x) H_l(y),$$

where  $m$  is the smallest integer for which there exists a non-zero Hermite coefficient  $a_{kl}$  with  $k + l = m$ .

**Definition 2.2.** The *Hermite rank* of a function  $h \in L^2(\mathbb{R}^2, \mathcal{N})$  is defined as

$$m = \inf\{k + l \mid k, l \geq 0, a_{kl} \neq 0\},$$

where  $a_{kl}$  is the coefficient in the Hermite expansion (12).

For an alternative approach and different notions of a two-dimensional Hermite rank, see the work by Lévy-Leduc and Taqqu (2014).

**Definition 2.3.** Let  $\xi, \eta \sim \mathcal{N}(0, 1)$  be two independent standard normal random variables. We define

$$\mathcal{G}^1(\mathbb{R}^2, \mathcal{N}) = \{G : \mathbb{R}^2 \rightarrow \mathbb{R} \text{ integrable} \mid E[G(\xi, \eta)] = 0\} \subset L^1(\mathbb{R}^2, \mathcal{N}),$$

the class of (with respect to the standard normal measure) centered and integrable functions on  $\mathbb{R}^2$ , and

$$\mathcal{G}^2(\mathbb{R}^2, \mathcal{N}) = \{G : \mathbb{R}^2 \rightarrow \mathbb{R} \text{ measurable} \mid E[G(\xi, \eta)] = 0, E[G^2(\xi, \eta)] = 1\} \subset L^2(\mathbb{R}^2, \mathcal{N}),$$

the class of (with respect to the standard normal measure) centered, normalized and square-integrable functions on  $\mathbb{R}^2$ . Analogously, we define the class  $\mathcal{G}^2(\mathbb{R}, \mathcal{N})$ .

Any function  $G : \mathbb{R}^2 \rightarrow \mathbb{R}$  which is measurable with mean zero and finite variance under standard normal measure can be normalized by dividing the standard deviation, so it can be considered as a function in  $\mathcal{G}^2 = \mathcal{G}^2(\mathbb{R}^2, \mathcal{N})$ .

**Theorem 2.1.** Let  $(\xi_i)_{i \geq 1}$  be a stationary Gaussian process with mean 0, variance 1 and covariances (2). Let  $Dm < 1$  and let  $h \in \mathcal{G}^2(\mathbb{R}^2, \mathcal{N})$  be a function with Hermite rank  $m$  whose Hermite coefficients satisfy

$$(15) \quad \sum_{k, l} \frac{|a_{kl}|}{\sqrt{k! l!}} < \infty.$$

Then as  $n \rightarrow \infty$

$$\frac{1}{d'_n n} \left| \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n h(\xi_i, \xi_j) - \sum_{\substack{k,l: \\ k+l=m}} \frac{a_{kl}}{k! l!} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n H_k(\xi_i) H_l(\xi_j) \right| \xrightarrow{L^1} 0$$

uniformly in  $\lambda \in [0, 1]$  and

(16)

$$\left( \frac{1}{d'_n n} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n h(\xi_i, \xi_j) \right)_{0 \leq \lambda \leq 1} \xrightarrow{\mathcal{D}} \left( \sum_{\substack{k,l: \\ k+l=m}} \frac{a_{kl}}{k! l!} (c_k c_l)^{1/2} Z_k(\lambda) (Z_l(1) - Z_l(\lambda)) \right)_{0 \leq \lambda \leq 1}$$

in  $D[0, 1]$ , where

$$(17) \quad d_n'^2 = n^{2-mD} L(n)^m$$

and the  $(Z_k(\lambda))_{\lambda \geq 0}$ ,  $k = 0, \dots, m$ , are dependent processes which can be expressed as  $k$ -fold Wiener-Itô-Integrals,

$$(18) \quad Z_k(\lambda) = K^{-k/2} c_k^{-1/2} \int'_{\mathbb{R}^k} \frac{e^{i\lambda \sum_{j=1}^k x_j} - 1}{i \sum_{j=1}^k x_j} \left( \prod_{j=1}^k |x_j|^{(D-1)/2} \right) dW(x_1) \dots dW(x_k),$$

where  $c_k$  is defined in (10) and where  $K = \int_{\mathbb{R}} e^{ix} |x|^{D-1} dx = 2\Gamma(D) \cos(D\pi/2)$ .

**Remark.** (i) Formula (18) denotes the multiple Wiener-Itô integral with respect to the random spectral measure  $W$  of the white-noise process, where  $\int'$  means that the domain of integration excludes the hyperdiagonals  $\{x_i = \pm x_j, i \neq j\}$ , see Major (1981b) or also Dehling and Taqqu (1989, p. 1769). The constant of proportionality  $c_m$  ensures that  $E[Z_m(1)]^2 = 1$ . Taqqu (1979) and Pipiras and Taqqu (2014, Chap. 3.2) give another representation.

(ii) The process  $(Z_m(\lambda))_{\lambda \geq 0}$  in (18) is called Hermite process of order  $m$ . For  $m = 1$ , this is a standard fractional Brownian motion. When  $m \geq 2$ , the process  $(Z_m(\lambda))_{\lambda \geq 0}$  is non-Gaussian.

(iii) The process  $(Z_m(\lambda))_{\lambda \geq 0}$  is self-similar with parameter

$$H = 1 - \frac{Dm}{2} \in \left( \frac{1}{2}, 1 \right),$$

i.e.  $(Z_m(c\lambda))_{\lambda \geq 0}$  and  $c^H (Z_m(\lambda))_{\lambda \geq 0}$  have the same finite-dimensional distributions for all constants  $c > 0$ .

(iv) Conditions of the type (15) are not uncommon in the study of U-statistic of dependent data. E.g., one finds such conditions in the recent papers by Neumann and Leucht (2013) and Denker and Gordin (2013).

(v) The scaling factor (17) differs slightly from the usual scaling (9) in that it does not include the normalizing constant  $c_m$ . This is caused by the fact that the limit now is a linear combination of two possibly different Hermite processes  $Z_k, Z_l$  and thus the associated factors  $c_k, c_l$  cannot be divided out and must remain inside the sum of the right-hand side of (16).

For the most interesting and simple case  $m = 1$ , we can give a handy explicit representation of the limit (16), because then  $Z_1(\lambda)$  is standard fractional Brownian motion  $B_H(\lambda)$  with  $H = 1 - D/2$ .

**Corollary.** *If the Hermite rank of  $h(x, y)$  is  $m = 1$ , the statement of Theorem 2.1 simplifies to*

$$(19) \quad \frac{1}{d'_n n} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n h(\xi_i, \xi_j) \xrightarrow{\mathcal{D}} \sqrt{c_1} (a_{1,0}(1-\lambda)B_H(\lambda) + a_{0,1}\lambda(B_H(1) - B_H(\lambda))),$$

where  $B_H(\lambda)$  is fractional Brownian motion with parameter  $H = 1 - D/2$ .

*Proof.* By Theorem 2.1, the limit is

$$\begin{aligned} & \sum_{k+l=1} \frac{a_{kl}}{k! l!} \sqrt{c_k c_l} Z_k(\lambda) (Z_l(1) - Z_l(\lambda)) \\ &= a_{1,0} \sqrt{c_1} Z_1(\lambda) (Z_0(1) - Z_0(\lambda)) + a_{0,1} \sqrt{c_1} Z_0(\lambda) (Z_1(1) - Z_1(\lambda)) \\ &= a_{1,0} \sqrt{c_1} (1 - \lambda) B_H(\lambda) + a_{0,1} \sqrt{c_1} \lambda (B_H(1) - B_H(\lambda)) \end{aligned}$$

with  $c_1 = 2/((1-D)(2-D))$ , because  $Z_0(t) = t$  and  $Z_1(t) = B_H(t)$ .  $\square$

In Section 4 we illustrate this by some examples. Since the approach is subject to technical restrictions which are sometimes difficult to check, we will develop some easily verifiable criteria for it in Section 3. Unfortunately, the technical restrictions are not satisfied by some special kernels like the Wilcoxon kernel  $h(x, y) = I_{\{x \leq y\}}$ . Thus, in Section 2.3, we present an approach using an empirical process representation of the two-sample U-statistic process.

**2.3. An approach via empirical processes.** Our second approach to the study of the asymptotic distribution of two-sample U-statistic processes uses a representation of  $U_n(\lambda)$  as a functional of the empirical distribution function. This approach has been used earlier for one-sample U-statistics by Dehling and Taqqu (1991); see also Beutner and Zähle (2012) for some recent extensions. Dehling and Taqqu (1989) have proved a limit theorem for the two-parameter empirical process  $(F_{[\lambda n]}(x) - F(x))_{x \in [-\infty, \infty], \lambda \in [0, 1]}$ , where

$$F_k(x) = \frac{1}{k} \sum_{i=1}^k I_{\{G(\xi_i) \leq x\}}$$

denotes the empirical distribution function (e.d.f.) of the first  $k$  observations  $G(\xi_1), \dots, G(\xi_k)$  and  $F$  denotes the cumulative distribution function (c.d.f.) of the  $G(\xi_i)$ . They consider the Hermite expansion

$$I_{\{G(\xi_i) \leq x\}} - F(x) = \sum_{k=1}^{\infty} \frac{J_k(x)}{k!} H_k(\xi_i),$$

where  $H_k$  again denotes the  $k$ -th Hermite polynomial and  $J_k(x)$  is the  $k$ -th Hermite coefficient in this expansion,

$$(20) \quad J_k(x) = E [I_{\{G(\xi) \leq x\}} H_k(\xi)] = (2\pi)^{-1/2} \int_{-\infty}^{\infty} I_{\{G(s) \leq x\}} H_k(s) e^{-s^2/2} ds$$

with  $\xi \sim \mathcal{N}(0, 1)$ .

**Definition 2.4** (Hermite rank of class of functions). We define the *Hermite rank of the class of functions*  $\{I_{\{G(\xi_i) \leq x\}} - F(x), x \in \mathbb{R}\}$  by

$$(21) \quad m = \min\{k \geq 1 : J_k(x) \neq 0 \text{ for some } x \in \mathbb{R}\}.$$

Now, we can state the second main result of the present paper.

**Theorem 2.2.** *Let  $(\xi_i)_{i \geq 1}$  be a stationary Gaussian process with mean zero, variance 1 and auto-covariance function as in (2). Moreover, let  $G : \mathbb{R} \rightarrow \mathbb{R}$  be a measurable function with  $E[G(\xi_i)] = 0$ , and define*

$$X_k = G(\xi_k).$$

*Assume that  $X_k$  has a continuous distribution function  $F$ . Let  $m$  denote the Hermite rank of the class of functions  $I_{\{G(\xi_i) \leq x\}} - F(x)$ ,  $x \in \mathbb{R}$ , and assume that  $mD < 1$ , where  $D$  is the exponent in (2). Moreover, define*

$$(22) \quad \tilde{h}(x) := \int h(x, y) dF(y),$$

*and assume that for some constant  $c \in (0, \infty)$ ,*

$$(23) \quad \|h(\cdot, y)\|_{TV} \leq c$$

$$(24) \quad \|h(x, \cdot)\|_{TV} \leq c.$$

*Then*

$$\left( \frac{1}{n d_n} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n \left( h(X_i, X_j) - \iint h(x, y) dF(x) dF(y) \right) \right)_{0 \leq \lambda \leq 1}$$

*converges in distribution towards the process*

$$(25) \quad \left( -(1-\lambda)Z(\lambda) \int J(x) d\tilde{h}(x) - \lambda(Z(1) - Z(\lambda)) \int \left( \int J(y) dh(x, y)(y) \right) dF(x) \right)_{0 \leq \lambda \leq 1}.$$

*Here  $Z(\lambda) = Z_m(\lambda)/m!$ , where  $(Z_m(\lambda))_{\lambda \geq 0}$  denotes the  $m$ -th order Hermite process defined in (18), and  $J(x) = J_m(x)$ .*

### 3. SUMMABILITY CRITERIA FOR HERMITE COEFFICIENTS

For most kernels, the summability condition (15) is not easily verified, as the Hermite coefficients cannot be explicitly calculated. In this section, we will derive alternative criteria that imply (15). Recall that the Fourier transform of a function  $f \in L^2(\mathbb{R}^d)$  is defined by

$$(26) \quad \mathcal{F}(f)(\xi) = \hat{f}(\xi) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} f(x) e^{-ix \cdot \xi} dx.$$

**Proposition 3.1.** *Let  $h \in L^2(\mathbb{R}^2)$ . Then condition (15) is fulfilled if*

$$(27) \quad \int |\mathcal{F}(h)(s, t)| (1 + s^2)(1 + t^2) ds dt < \infty.$$

*A sufficient condition for (27) to hold is that  $\frac{\partial^8}{\partial x^4 \partial y^4} h(x, y)$  exists and is in  $L^1(\mathbb{R}^2)$ .*

*Proof.* By the Plancherel theorem we can write the Hermite coefficients in the following way:

$$\begin{aligned} a_{kl} &= \frac{1}{2\pi} \iint_{\mathbb{R}^2} h(x, y) H_k(x) H_l(y) e^{-(x^2+y^2)/2} dx dy \\ &= \frac{1}{2\pi} \iint_{\mathbb{R}^2} \hat{h}(s, t) g(s, t) ds dt, \end{aligned}$$



where  $g(s, t) := \mathcal{F}(H_k(x)H_l(y)e^{-(x^2+y^2)/2})$ . In order to give an explicit representation of  $g(s, t)$ , we will apply some well-known properties of Fourier transforms and Hermite polynomials, namely

$$\begin{aligned}\mathcal{F}\left(e^{-(x^2+y^2)/2}\right) &= e^{-(s^2+t^2)/2} \\ \mathcal{F}\left(\frac{\partial^{k+l}}{\partial x^k \partial y^l} f(x, y)\right) &= i^{k+l} s^k t^l \hat{f}(s, t) \\ H_k(x)H_l(y)e^{-(x^2+y^2)/2} &= (-1)^{k+l} \frac{\partial^{k+l}}{\partial x^k \partial y^l} e^{-(x^2+y^2)/2}.\end{aligned}$$

With these formulae we can write

$$a_{kl} = \frac{1}{2\pi} \iint_{\mathbb{R}^2} \hat{h}(s, t) (-i)^{k+l} s^k t^l e^{-(s^2+t^2)/2} ds dt,$$

and hence we obtain

$$\begin{aligned}\sum_{k,l} \frac{|a_{kl}|}{\sqrt{k!} \sqrt{l!}} &= \frac{1}{2\pi} \sum_{k,l} \left| \iint_{\mathbb{R}^2} \hat{h}(s, t) (-i)^{k+l} \frac{s^k t^l}{\sqrt{k!} \sqrt{l!}} e^{-(s^2+t^2)/2} ds dt \right| \\ &\leq \frac{1}{2\pi} \sum_{k,l} \iint_{\mathbb{R}^2} |\hat{h}(s, t)| \frac{|s|^k |t|^l}{\sqrt{k!} \sqrt{l!}} e^{-(s^2+t^2)/2} ds dt \\ (28) \quad &= \frac{1}{2\pi} \iint_{\mathbb{R}^2} |\hat{h}(s, t)| \left( \sum_{k=1}^{\infty} \frac{|s|^k}{\sqrt{k!}} \right) \left( \sum_{l=1}^{\infty} \frac{|t|^l}{\sqrt{l!}} \right) e^{-(s^2+t^2)/2} ds dt.\end{aligned}$$

Now, we will bound the series  $\sum_{k=1}^{\infty} \frac{s^k}{\sqrt{k!}}$ , for  $s \geq 0$ . For notational convenience, we write  $(k/2)! := \Gamma(k/2 + 1)$ . Using Stirling's formula for the Gamma function, we then obtain  $(k/2)!/\sqrt{k!} \sim C 2^{-k/2} k^{1/4} \leq C k 2^{-k/2}$ , and thus

$$\sum_{k=1}^{\infty} \frac{s^k}{\sqrt{k!}} \leq C \sum_{k=1}^{\infty} k \frac{s^k 2^{-k/2}}{(k/2)!} = C \sum_{k=1}^{\infty} k \frac{(s^2/2)^{k/2}}{(k/2)!}.$$

Now we use the series expansion

$$(29) \quad \sum_{k=1}^{\infty} \frac{x^{k/2} k/2}{(k/2)!} = \frac{\sqrt{x}}{\sqrt{\pi}} + x e^x (1 + \operatorname{erf}(\sqrt{x})),$$

where  $\operatorname{erf}(x) = 2/\sqrt{\pi} \int_0^x e^{-t^2} dt$  denotes the Gaussian error function. The identity (29) follows from the series expansion  $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} e^{-z^2} \sum_{k=0}^{\infty} \frac{2^k}{1 \cdot 3 \cdots (2k+1)} z^{2k+1}$ ; see Gradshteyn and Ryzhik (1980), p. 931. Using the fact that the Gaussian error function is bounded by 1, we obtain

$$\sum_{k=1}^{\infty} \frac{s^k}{\sqrt{k!}} \leq C \left( s + s^2 e^{s^2/2} \right),$$

and hence

$$\begin{aligned}\sum_{k,l} \frac{|a_{kl}|}{\sqrt{k!} \sqrt{l!}} &\leq C \iint_{\mathbb{R}^2} |\hat{h}(s, t)| \left( s e^{-s^2/2} + s^2 \right) \left( t e^{-t^2/2} + t^2 \right) ds dt \\ &\leq C \iint_{\mathbb{R}^2} |\hat{h}(s, t)| (1 + s^2) (1 + t^2) ds dt.\end{aligned}$$

Since  $\mathcal{F}(\frac{\partial^8}{\partial^4 x \partial^4 y} h(x, y)) = s^4 t^4 \mathcal{F}(h)(s, t)$ , and since  $\frac{\partial^8}{\partial^4 x \partial^4 y} h(x, y) \in L^1$  by assumption, we obtain from the Riemann-Lebesgue lemma that  $s^4 t^4 \mathcal{F}(h)(s, t) \rightarrow 0$  as  $s, t \rightarrow \infty$ . Thus  $\mathcal{F}(h)(s, t) = o(\frac{1}{s^4 t^4})$ , and hence the integral (27) is finite.  $\square$

**Example.** Any function  $h \in L^1(\mathbb{R}^2, \lambda)$  whose derivatives of order 8 are integrable satisfies the summability condition (15), for instance:

(i) a (normalized) Hermite function

$$\tilde{h}_{kl}(x, y) = \frac{1}{\sqrt{2^{k+l} k! l! \pi}} H_{kl}(x, y) e^{-(x^2+y^2)/2}$$

(ii) a Gaussian function

$$g(x, y) = a \exp \left\{ -b \cdot \left( \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \right)^t \Sigma^{-1} \left( \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \right) \right\}$$

with  $a, b, \mu_1, \mu_2 \in \mathbb{R}$  and  $\Sigma \in \mathbb{R}^{2 \times 2}$  a symmetric positive-definite matrix

(iii) a smooth function with bounded support like the bump function

$$f(x, y) = \begin{cases} e^{\frac{-1}{1-x^2}} e^{\frac{-1}{1-y^2}} & |x|, |y| < 1 \\ 0 & \text{else} \end{cases}$$

## 4. EXAMPLES

### 4.1. Examples related to the Hermite expansion approach.

4.1.1. *CUSUM statistic.* The kernel  $h(x, y) = x - y$  leads to the *CUSUM statistic*

$$U_{C,n}(\lambda) = \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n (\xi_i - \xi_j) = [\lambda n](n - [\lambda n]) \left( \frac{1}{[\lambda n]} \sum_{i=1}^{[\lambda n]} X_i - \frac{1}{n - [\lambda n]} \sum_{i=[\lambda n]+1}^n X_i \right),$$

a scaled difference of the means of the first and the second part of the sample. The kernel  $h(x, y) = x - y$  is of course in  $L^2(\mathbb{R}^2, \mathcal{N})$  and its Hermite expansion can be read off without calculating:

$$h(x, y) = x - y = \frac{a_{1,0}}{1!0!} H_1(x) + \frac{a_{0,1}}{0!1!} H_1(y),$$

so its Hermite coefficients are

$$a_{kl} = \begin{cases} 1 & k = 1, l = 0 \\ -1 & k = 0, l = 1 \\ 0 & \text{else} \end{cases},$$

and condition (15) is trivially fulfilled. Thus, the Corollary to Theorem 2.1 yields

$$\frac{1}{d_n'' n} U_{C,n}(\lambda) \xrightarrow{\mathcal{D}} \sqrt{c_1} ((1 - \lambda) B_H(\lambda) - \lambda (B_H(1) - B_H(\lambda))),$$

so we have reproduced the result of Horváth and Kokoszka (1997).

So far, we have considered Gaussian observations  $\xi_1, \dots, \xi_n$ . As mentioned in the introduction, all results can be extended to general data  $G(\xi_1), \dots, G(\xi_n)$ , where  $G \in \mathcal{G}^2(\mathbb{R}, \mathcal{N}) \subset L^2(\mathbb{R}, \mathcal{N})$  is a transformation like the quantile transformations in the papers of Dehling,

Rooch and Taqqu (2013a,b), by considering the kernel  $h(G(x), G(y))$  instead of  $h(x, y)$ . This is what we will do now exemplarily. The Hermite coefficients of the function  $h(G(x), G(y))$  are

$$\begin{aligned} a_{kl} &= \iint_{\mathbb{R}^2} (G(x) - G(y)) H_k(x) H_l(y) d\Phi(x) d\Phi(y) \\ &= \int_{\mathbb{R}} G(x) H_k(x) d\Phi(x) \cdot \int_{\mathbb{R}} H_l(y) d\Phi(y) - \int_{\mathbb{R}} G(y) H_l(y) d\Phi(y) \cdot \int_{\mathbb{R}} H_k(x) d\Phi(x) \\ &= \begin{cases} 0 & \text{if } k, l \neq 0 \\ -a_l & \text{if } k = 0, l \neq 0, \\ a_k & \text{if } k \neq 0, l = 0 \end{cases} \end{aligned}$$

where  $a_p = E[G(\xi)H_p(\xi)]$  is the  $p$ -th Hermite coefficient of  $G$ . Thus for  $G : \mathbb{R} \rightarrow \mathbb{R}$ ,  $G \in \mathcal{G}^2(\mathbb{R}, \mathcal{N})$  and  $h(x, y) = x - y$ , the summability condition (15) turns into a similar condition on the transformation  $G$ :

$$\sum_{k,l} \frac{|a_{kl}|}{\sqrt{k!l!}} = 2 \sum_{k=1}^{\infty} \frac{|a_k|}{\sqrt{k!}} < \infty$$

4.1.2. *The Wilcoxon statistic.* The kernel  $h(x, y) = I_{\{x \leq y\}}$  yields the well-known *Mann-Whitney-Wilcoxon statistic*

$$U_{W,n}(\lambda) = \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n I_{\{\xi_i \leq \xi_j\}}.$$

We will now show that this kernel does not fulfil the summability condition (15). Nevertheless, if we ignore this, the above theorems reproduce the results from Dehling, Rooch and Taqqu (2013a) which suggests that condition (15) may be too strong and Theorem 2.1 may hold under milder assumptions.

That  $h(x, y) = I_{\{x \leq y\}}$  does not satisfy (15) is neither intuitively visible nor enjoyable to verify. One can show, using integration by parts and identities for Hermite polynomials and the Gamma function, that the Hermite coefficients of  $h(x, y) = I_{\{x \leq y\}}$  are

$$(30) \quad a_{k,l} = \begin{cases} \frac{(-1)^{\frac{l+3k-1}{2}}}{2\pi} \Gamma\left(\frac{l+k}{2}\right) & l+k \text{ odd and positive} \\ 0 & l+k \text{ even and positive} \\ \frac{1}{2} & l=k=0 \end{cases}$$

In order to show that  $\sum_{k,l=1}^{\infty} |a_{k,l}|/\sqrt{k!l!}$  diverges, it is enough to consider the first odd diagonal where  $l = k + 1$ , because there we have already with Sterling's approximation

$$\frac{|a_{k,l}|}{\sqrt{k!l!}} \sim \frac{(2k-1)^k e}{2^k (k+1)^{k/2+3/4} k^{k/2+1/4}} = \frac{(1 - \frac{1}{2k})^k \frac{1}{k} e}{(1 + \frac{1}{k})^{k/2} (1 + \frac{1}{k})^{3/4}} \sim \frac{1}{k}.$$

Let us for a moment ignore that the Wilcoxon kernel does not fulfill the summability condition (15), which may be too rigorous anyway, and apply Theorem 2.1. To this end, we use

(30) or calculate the first Hermite coefficients manually:

$$\begin{aligned} a_{0,0} &= \iint_{\{x \leq y\}} H_0(x) H_0(y) \varphi(x) \varphi(y) dx dy = \iint_{\{x \leq y\}} \varphi(x) \varphi(y) dx dy = \frac{1}{2} \\ a_{1,0} &= \iint_{\{x \leq y\}} x \varphi(x) \varphi(y) dx dy = -\frac{1}{2\sqrt{\pi}} \\ a_{0,1} &= \iint_{\{x \leq y\}} y \varphi(x) \varphi(y) dx dy = \frac{1}{2\sqrt{\pi}} \end{aligned}$$

Since we formulated the theorem for centered kernels, we consider  $h(x, y) - E[h(\xi, \eta)] = I_{\{x \leq y\}} - 1/2$ , which has Hermite rank  $m = 1$ . So the Corollary to Theorem 2.1 states that

$$\begin{aligned} \frac{1}{n d'_n} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n I_{\{\xi_i \leq \xi_j\}} &\xrightarrow{\mathcal{D}} \sqrt{c_1} (a_{1,0}(1 - \lambda) B_H(\lambda) + a_{0,1} \lambda (B_H(1) - B_H(\lambda))) \\ &= \frac{\sqrt{c_1}}{2\sqrt{\pi}} (\lambda B_H(1) - B_H(\lambda)). \end{aligned}$$

Bearing in mind that  $\int_{\mathbb{R}} J_1(x) d\Phi(x) = -(2\sqrt{\pi})^{-1}$ , we have just reproduced the results of Dehling, Rooch and Taqqu (2013a) for the Gaussian case.

## 4.2. Examples related to the empirical process approach.

4.2.1. *The kernel  $h(x, y) = I_{\{x \leq y\}}$ .* Dehling, Rooch and Taqqu (2013a) have investigated the asymptotic distribution of the *Mann-Whitney-Wilcoxon statistic*

$$W_{[\lambda n], n} = \frac{1}{n d_n} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n \left( I_{\{X_i \leq X_j\}} - \frac{1}{2} \right).$$

We will now show that the conditions of Theorem 2.2 are satisfied, and that we obtain the same result as Dehling, Rooch and Taqqu (2013a). For fixed  $x$ , the kernel  $h(x, y) = 1_{\{x \leq y\}}$  is of bounded variation as a function of  $y$ , and the same holds for fixed  $y$ . In fact (23) and (24) are satisfied with  $c = 1$ . Moreover, we obtain

$$\begin{aligned} \tilde{h}(x) &= \int h(x, y) dF(y) = \int_x^\infty dF(y) = 1 - F(x) \\ \int J(x) d\tilde{h}(x) &= - \int J(x) dF(x) \\ \iint J(y) dh(x, y)(y) dF(x) &= \int J(x) dF(x). \end{aligned}$$

So Theorem 2.2 reproduces the result of Dehling, Rooch and Taqqu (2013a).

4.2.2. *A class of kernels for robust change-point detection.* By Theorem 2.6 in Huber (1981), an M-estimator of location is robust iff the score function  $\Psi$  is bounded. As score functions are typically either nondecreasing or redescending, they have finite total variation. Examples include the score functions introduced by Andrews, Hampel, Huber, Tukey; see Huber (1981). The hypothesis of no change corresponds to the location of the differences  $X_i - X_j$ ,

$1 \leq i \leq k$ ,  $k+1 \leq j \leq n$  being 0. This motivates the following class of robust change point statistics: We take the maximum of the two-sample  $U$ -statistic process with kernel

$$h(x, y) := \Psi(x - y)$$

where  $\Psi$  is a robust score function and hence of bounded total variation. Obviously, conditions (23) and (24) of Theorem 2.2 hold.

4.2.3. *The kernel  $h(x, y) = x - y$ .* This kernel leads to the classical CUSUM statistic

$$\begin{aligned} \frac{1}{n d_n} U_{\text{diff}, [\lambda n], n} &= \frac{1}{n d_n} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n (X_i - X_j) \\ &= \frac{[\lambda n](n - [\lambda n])}{n d_n} (\bar{X}_{[\lambda n]} - \bar{X}_{[\lambda n]+1, n}), \end{aligned}$$

where  $\bar{X}_{[\lambda n]}$  denotes the arithmetic mean of the first  $[\lambda n]$  observations and  $\bar{X}_{[\lambda n]+1, n}$  denotes the arithmetic mean of the last  $n - [\lambda n]$  observations. Here, the conditions of Theorem 2.2 are not met, since  $h$  is not of bounded total variation. Nevertheless, both integrals occurring in the limit (25) exist, and thus we can formally apply Theorem 2.2. In order to show this, we first note that

$$\tilde{h}(x) = \int_{\mathbb{R}} h(x, y) dF(y) = \int_{\mathbb{R}} (x - y) d\Phi(y) = x.$$

Moreover  $dh(x, y)(y) = d(x - y)(y) = -dy$ . Both integrals in the limit exist and have the same absolute value, namely

$$\left| \int_{\mathbb{R}} J(y) dh(x, y)(y) \right| = \int_{\mathbb{R}} \varphi(y) dy = 1.$$

So, if one ignored that some of the conditions are violated, Theorem 2.2 would state that  $U_{\text{diff}, [\lambda n], n}$  converges in distribution to the process

$$\begin{aligned} (1 - \lambda)Z(\lambda) \int \varphi(x) dx - \lambda(Z(1) - Z(\lambda)) \int \left( \int \varphi(y) dy \right) d\Phi(x) \\ = (1 - \lambda)Z(\lambda) - \lambda(Z(1) - Z(\lambda)) \\ = Z(\lambda) - \lambda Z(1), \end{aligned}$$

where  $Z(\lambda) = B_{1-D/2}(\lambda)$  denotes the standard fractional Brownian motion with Hurst parameter  $H = 1 - D/2$ . A rigorous proof of this result was obtained by Horváth and Kokoszka (1997).

## 5. PROOFS

5.1. **Proof of Theorem 2.1.** The expansion (14) of  $h(x, y)$  in Hermite polynomials converges to  $h$  in  $L^2(\mathbb{R}^2, \mathcal{N})$ . Thus, for independent standard normal random variables  $\xi, \eta$ , we obtain the series expansion

$$h(\xi, \eta) = \sum_{q=m}^{\infty} \sum_{k, l: k+l=q} \frac{a_{kl}}{k!l!} H_k(\xi) H_l(\eta).$$

In order to handle  $h(\xi_i, \xi_j)$ , we face the problem that any pair  $(\xi_i, \xi_j)$  is dependent and has a joint normal distribution with covariance matrix that is not the identity. So first we ensure

that the expansion (14) is nevertheless applicable in our situation. We show first that under condition (15), (14) converges almost surely to  $h(x, y)$ .

By the Cauchy-Schwarz inequality and (15), we obtain

$$\begin{aligned} E \left[ \sum_{k,l} \left| \frac{a_{kl}}{k!l!} H_k(\xi_i) H_l(\xi_j) \right| \right] &\leq \sum_{k,l} \frac{|a_{kl}|}{k!l!} E |H_k(\xi_i) H_l(\xi_j)| \\ &\leq \sum_{k,l} \frac{|a_{kl}|}{k!l!} \sqrt{E [H_k^2(\xi_i)] E [H_l^2(\xi_j)]} \\ &= \sum_{k,l} \frac{|a_{kl}|}{\sqrt{k!l!}} < \infty. \end{aligned}$$

Thus, the series  $\sum_{k,l} \frac{a_{kl}}{k!l!} H_k(\xi_i) H_l(\xi_j)$  is almost surely absolutely convergent, and the same holds for the series  $\sum_{k,l} \frac{a_{kl}}{k!l!} H_k(x) H_l(y)$ , with respect to any bivariate normal distribution. Since we have  $L^2$ -convergence to  $h(x, y)$ , with respect to the bivariate standard normal distribution, and since all nondegenerate normal distributions on  $\mathbb{R}^2$  are equivalent, the almost sure limit of  $\sum_{k,l} \frac{a_{kl}}{k!l!} H_k(\xi_i) H_l(\xi_j)$  is  $h(\xi_i, \xi_j)$ . Thus we have

$$h(\xi_i, \xi_j) = \sum_{\substack{k,l: \\ k+l \geq m}} \frac{a_{kl}}{k!l!} H_k(\xi_i) H_l(\xi_j)$$

and hence

$$\sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n h(\xi_i, \xi_j) = \sum_{k,l: k+l \geq m} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n \frac{a_{kl}}{k!l!} H_k(\xi_i) H_l(\xi_j).$$

Thus, we obtain

$$\begin{aligned} &\sup_{0 \leq \lambda \leq 1} \frac{1}{d'_n n} \left( \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n h(\xi_i, \xi_j) - \sum_{\substack{k,l: \\ k+l=m}} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n \frac{a_{kl}}{k!l!} H_k(\xi_i) H_l(\xi_j) \right) \\ &= \sup_{0 \leq \lambda \leq 1} \frac{1}{d'_n n} \sum_{\substack{k,l: \\ k+l \geq m+1}} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n \frac{a_{kl}}{k!l!} H_k(\xi_i) H_l(\xi_j). \end{aligned}$$

We will show that the term on the right-hand side converges in  $L^1$  to 0. Note first that the supremum here is in fact a maximum, since  $\lambda$  appears only in terms of the integer  $[\lambda n]$ , thus by setting  $b = [\lambda n]$  and by the fact that we need to have  $[\lambda n] \geq 1$  in order to have a two-sample statistic, we can replace  $\sup_{0 \leq \lambda \leq 1}$  by  $\max_{1 \leq b \leq n}$ . Using the inequality

$\max_b |f(b)g(b)| \leq \max_b |f(b)| \max_b |g(b)|$  and the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
 & E \left| \frac{1}{d'_n n} \sup_{0 \leq \lambda \leq 1} \sum_{\substack{k,l: \\ k+l \geq m+1}} \frac{a_{kl}}{k! l!} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n H_k(\xi_i) H_l(\xi_j) \right| \\
 & \leq E \left[ \frac{1}{d'_n n} \sum_{\substack{k,l: \\ k+l \geq m+1}} \frac{|a_{kl}|}{k! l!} \max_{1 \leq b \leq n} \left| \sum_{i=1}^b H_k(\xi_i) \right| \max_{1 \leq b \leq n} \left| \sum_{j=b+1}^n H_l(\xi_j) \right| \right] \\
 (31) \quad & \leq \frac{1}{d'_n n} \sum_{\substack{k,l: \\ k+l \geq m+1}} \frac{|a_{kl}|}{k! l!} \left( E \left[ \max_{1 \leq b \leq n} \left| \sum_{i=1}^b H_k(\xi_i) \right| \right]^2 E \left[ \max_{1 \leq b \leq n} \left| \sum_{j=b+1}^n H_l(\xi_j) \right| \right]^2 \right)^{1/2}.
 \end{aligned}$$

In order to show that this term converges to 0, we need bounds for the expectations of the squared maxima. The growth of the partial sum  $\sum_{i=1}^b H_k(\xi_i)$  is determined by the degree  $k$  of the Hermite polynomial and the size of the LRD parameter  $D \in (0, 1)$ : For  $Dk > 1$  we observe usual SRD behaviour, while for  $Dk < 1$  we observe a faster rate of growth, remember (8). First we consider the SRD case, that is  $Dk > 1$ . Here we have by (8)

$$E \left[ \sum_{i=1}^b H_k(\xi_i) \right]^2 \leq Ck! \cdot b,$$

and thus we obtain by stationarity and a maximal inequality of Móricz (1976, Theorem 3)

$$(32) \quad E \left[ \max_{1 \leq b \leq n} \left| \sum_{i=1}^b H_k(\xi_i) \right| \right]^2 \leq 4Ck! \cdot n(\log_2 n)^2.$$

Here we used the estimate  $\log_2(2n) \leq 2 \log_2 n$  for  $n \geq 2$ .

Now we turn to the LRD case, that is  $Dk < 1$ . Here we have by (8) and the simple estimate  $b^{2-Dk} \leq bn^{1-Dk}$  for all  $b \leq n$  (and we do not consider any other  $b$ )

$$E \left[ \sum_{i=1}^b H_k(\xi_i) \right]^2 \leq \tilde{C}(k)k! \cdot n^{1-Dk} \max_{1 \leq b \leq n} L^k(b) \cdot b,$$

and thus by the same inequality of Móricz (1976, Theorem 3)

$$(33) \quad E \left[ \max_{1 \leq b \leq n} \left| \sum_{i=1}^b H_k(\xi_i) \right| \right]^2 \leq 4\tilde{C}(k)k! \cdot n^{2-Dk} \max_{1 \leq b \leq n} L^k(b) \cdot (\log_2 n)^2.$$

Note that the same estimates hold in (31) for the sum that starts at  $b+1$  because for some  $b' \in \{1, \dots, n\}$

$$\max_{1 \leq b \leq n} \left| \sum_{j=b+1}^n H_l(\xi_j) \right| \stackrel{\mathcal{D}}{=} \max_{1 \leq b \leq n} \left| \sum_{j=1}^{n-b} H_l(\xi_j) \right| = \max_{1 \leq b' \leq n} \left| \sum_{j=1}^{b'} H_l(\xi_j) \right|$$

where  $\stackrel{\mathcal{D}}{=}$  denotes equality in distribution since the  $(\xi_i)_{i \geq 1}$  are stationary.

Now depending on the size of  $k$  and  $l$ , both sums in (31) can be SRD or LRD – and thus they can be bounded by (32) or by (33) –, such that we have to discriminate four cases.

Case 1:  $k, l < 1/D$ , i.e. both sums are LRD. In this case, (31) is bounded by

$$\begin{aligned} & \frac{1}{n^{2-Dm/2} L^{m/2}(n)} \sum_{\substack{k+l \geq m+1 \\ k, l < 1/D}} \frac{|a_{kl}|}{k! l!} \left( \tilde{C}(k) \sqrt{k!} n^{1-Dk/2} \max_{1 \leq b \leq n} L^{k/2}(b) \log_2 n \right. \\ & \quad \left. \cdot \tilde{C}(l) \sqrt{l!} n^{1-Dl/2} \max_{1 \leq b \leq n} L^{l/2}(b) \log_2 n \right) \\ & \leq \sum_{\substack{k+l \geq m+1 \\ k, l < 1/D}} \frac{|a_{kl}|}{\sqrt{k! l!}} \left( \tilde{C}(k) \tilde{C}(l) \right. \\ & \quad \left. \cdot n^{\frac{D}{2}(m-(k+l))} \max_{1 \leq b \leq n} L^{k/2}(b) \max_{1 \leq b \leq n} L^{l/2}(b) L^{-m/2}(n) (\log_2 n)^2 \right) \end{aligned}$$

Now  $n^{\frac{D}{2}(m-(k+l))} = n^{-\varepsilon}$  for some  $\varepsilon > 0$ , and  $L^{-m/2}(n)$  and  $\log_2^2 n$  are  $o(n^\varepsilon)$  for any  $\varepsilon > 0$ . We will immediately show that also  $\max_{1 \leq b \leq n} L^{k/2}(b)$  is  $o(n^\varepsilon)$  for any  $\varepsilon > 0$  and  $k \in \mathbb{N}$ . Because the summation over  $k, l$  is only finite, the sum on the right-hand side is finite, and thus the right-hand side converges to 0.

Now we show that  $\max_{1 \leq b \leq n} L^{k/2}(b)$  is  $o(n^\varepsilon)$  for any  $\varepsilon > 0$  and  $k \in \mathbb{N}$ . When  $L$  is slowly varying,  $L^{k/2}(x)$  is it as well. So we need to consider

$$\begin{aligned} \max_{1 \leq b \leq n} \frac{L(b)}{n^\varepsilon} & \leq \max_{1 \leq b \leq \sqrt{n}} \frac{L(b)}{\sqrt{n}^\varepsilon \sqrt{n}^\varepsilon} + \max_{\sqrt{n} \leq b \leq n} \frac{L(b)}{n^\varepsilon} \\ & \leq \frac{1}{\sqrt{n}^\varepsilon} \max_{1 \leq b \leq \sqrt{n}} \frac{L(b)}{b^\varepsilon} + \max_{\sqrt{n} \leq b \leq n} \frac{L(b)}{b^\varepsilon}, \end{aligned}$$

and since  $L(b)/b^\varepsilon \rightarrow 0$  as  $b \rightarrow \infty$ ,  $\max_{1 \leq b \leq \sqrt{n}} \frac{L(b)}{b^\varepsilon}$  is bounded and  $\max_{\sqrt{n} \leq b \leq n} \frac{L(b)}{b^\varepsilon}$  converges to 0.

Case 2:  $k < 1/D$  and  $l > 1/D$ , i.e. the sum over  $i$  is LRD and the sum over  $j$  is SRD. In this case, (31) is bounded by

$$\begin{aligned} & \frac{1}{n^{2-Dm/2} L^{m/2}(n)} \sum_{\substack{k+l \geq m+1 \\ k < 1/D, l > 1/D}} \frac{|a_{kl}|}{k! l!} \left( C(k) \sqrt{k!} n^{1-Dk/2} \max_{1 \leq b \leq n} L^{k/2}(b) \log_2 n \right. \\ & \quad \left. \cdot \sqrt{l!} \sqrt{n} \log_2 n \right) \\ & \leq \sum_{\substack{k+l \geq m+1 \\ k < 1/D, l > 1/D}} \frac{|a_{kl}|}{\sqrt{k! l!}} \left( C(k) n^{-\frac{1}{2} + \frac{Dm}{2} - \frac{Dk}{2}} \max_{1 \leq b \leq n} L^{k/2}(b) L^{-m/2}(n) (\log_2 n)^2 \right) \end{aligned}$$

Here, we have summed up some constants in order to keep the expression simple. Now  $n^{-\frac{1}{2} + \frac{Dm}{2} - \frac{Dk}{2}} = n^{-\varepsilon}$  for a  $\varepsilon > 0$ , because  $\frac{Dm}{2}, \frac{Dk}{2} \in (0, \frac{1}{2})$ .  $\max_{1 \leq b \leq n} L^{k/2}(b)$ ,  $L^{-m/2}(n)$  and



$\log_2^2 n$  are  $o(n^\varepsilon)$  for any  $\varepsilon > 0$  as above, and the sum on the right hand side is finite, because of (15) and since the summation over  $k$  is only finite.

Case 3:  $k > 1/D$  and  $l < 1/D$ , i.e. the sum over  $i$  is SRD and the sum over  $j$  is LRD. In this case, (31) converges to 0 by the same arguments.

Case 4:  $k, l > 1/D$ , i.e. both sums are SRD. In this case, (31) is bounded by

$$\begin{aligned} & \frac{1}{n^{2-Dm/2} L^{m/2}(n)} \sum_{\substack{k+l \geq m+1 \\ k, l > 1/D}} \frac{|a_{kl}|}{k! l!} \left( C \sqrt{k!} \sqrt{n} \log_2 n \cdot \sqrt{l!} \sqrt{n} \log_2 n \right) \\ & \leq C \sum_{\substack{k+l \geq m+1 \\ k, l > 1/D}} \frac{|a_{kl}|}{\sqrt{k! l!}} \left( n^{-1+Dm/2} L^{-m/2}(n) \log_2^2 n \right) \end{aligned}$$

Now  $n^{-1+Dm/2} = n^{-\varepsilon}$  for a  $\varepsilon > 0$ , because  $\frac{Dm}{2} \in (0, \frac{1}{2})$ .  $L^{-m/2}(n)$  and  $\log_2^2 n$  are  $o(n^\varepsilon)$  for any  $\varepsilon > 0$  as above, and the sum on the right hand side is finite, because of (15).

Thus, in all four cases, (31) converges to 0, and thus the first statement of the theorem is proved.

For the proof of the second statement of the theorem, we apply the multivariate non-CLT for LRD processes of Taqqu and Bai (2012), which states that for any integer  $m < 1/D$

$$\left( \frac{1}{d_n(1)} \sum_{i=1}^{[\lambda_1 n]} H_1(\xi_i), \frac{1}{d_n(2)} \sum_{i=1}^{[\lambda_2 n]} H_2(\xi_i), \dots, \frac{1}{d_n(m)} \sum_{i=1}^{[\lambda_m n]} H_m(\xi_i) \right)$$

converges in distribution to the  $m$ -dimensional process

$$(Z_1(\lambda_1), Z_2(\lambda_2), \dots, Z_m(\lambda_m))$$

in  $D[0, 1]^m$ , where  $d_n^2(k)$  is the usual scaling for the partial sum of the  $k$ -th Hermite polynomial, as defined in (9), and  $Z_k$  denotes the  $k$ -th order Hermite process, as defined in (18), driven by the same Wiener process  $W$ .

We now consider the cases where  $k+l = m$ . By the multivariate non-CLT and the continuous mapping theorem, we obtain

$$\begin{aligned} & \frac{1}{d'_n n} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n H_k(\xi_i) H_l(\xi_j) \\ & = n^{-2+Dm/2} L(n)^{-m/2} \sum_{i=1}^{[\lambda n]} H_k(\xi_i) \left( \sum_{j=1}^n H_l(\xi_j) - \sum_{j=1}^{[\lambda n]} H_l(\xi_j) \right) \\ & \xrightarrow{\mathcal{D}} c_k^{1/2} Z_k(\lambda) \cdot c_l^{1/2} (Z_l(1) - Z_l(\lambda)) \end{aligned}$$

uniformly in  $\lambda \in [0, 1]$ . □

**5.2. Proof of Theorem 2.2.** We will express the two-sample U-statistic process as a functional of the empirical process of the observations  $(X_i)_{i \geq 1}$  and apply the empirical process non-CLT of Dehling and Taqqu (1989), which states that for  $Dm < 1$ , as  $n \rightarrow \infty$

$$(34) \quad \left( \frac{[\lambda n]}{d_n} (F_{[\lambda n]}(x) - F(x)) \right)_{x \in [-\infty, \infty], \lambda \in [0, 1]} \xrightarrow{\mathcal{D}} \left( \frac{J_m(x)}{m!} Z_m(\lambda) \right)_{x \in [-\infty, \infty], \lambda \in [0, 1]}.$$

We denote by  $F_{k,n}$  the empirical distribution function of the observations  $X_k, \dots, X_n$ , i.e.

$$F_{k,n}(x) = \frac{1}{n-k+1} \sum_{i=k}^n 1_{\{X_i \leq x\}}.$$

Dehling, Rooch and Taqqu (2013a) have shown that, applying the Dudley-Wichura version of Skorohod's representation theorem (Shorack and Wellner, 1986, Th. 2.3.4) to (34), we may assume without loss of generality that

$$(35) \quad \sup_{\lambda, x} |d_n^{-1}[n\lambda](F_{[\lambda n]}(x) - F(x)) - J(x)Z(\lambda)| \longrightarrow 0 \quad a.s.$$

$$(36) \quad \sup_{\lambda, x} |d_n^{-1}(n - [n\lambda])(F_{[n\lambda]+1,n}(x) - F(x)) - J(x)(Z(1) - Z(\lambda))| \longrightarrow 0 \quad a.s. .$$

Next, we obtain some auxiliary results concerning the kernel  $h$ , and the integrals that occur in the limit (25). First, note that (23) and (24) imply that the kernel  $h$  is bounded. Moreover, (23) implies that  $\tilde{h}$  is of bounded total variation, since for any  $x_0 < x_1 < \dots < x_k$

$$\begin{aligned} \sum_{i=1}^k |\tilde{h}(x_i) - \tilde{h}(x_{i-1})| &= \sum_{i=1}^k \left| \int (h(x_i, y) - h(x_{i-1}, y)) dF(y) \right| \\ &\leq \int \sum_{i=1}^k (|h(x_i, y) - h(x_{i-1}, y)|) dF(y) \leq c. \end{aligned}$$

Thus  $\|\tilde{h}\|_{TV} \leq c$  and, as a consequence,  $\tilde{h}$  is bounded. By definition (20), the function  $J(x)$  is bounded, and thus

$$(37) \quad \left| \int J(x) d\tilde{h}(x) \right| \leq \|J\|_{\infty} c,$$

where  $\|J\|_{\infty}$  denotes the supremum norm of  $J$ . In the same way, we obtain

$$(38) \quad \int \left( \int J(y) dh(x, y)(y) \right) dF(x) \leq \int \|J\|_{\infty} c dF(x) \leq \|J\|_{\infty} c.$$

We now express the two-sample U-statistic as a functional of the empirical process, and obtain

$$\begin{aligned} &\sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n \left( h(X_i, X_j) - \iint h(x, y) dF(x) dF(y) \right) \\ &= [\lambda n](n - [n\lambda]) \left\{ \int \left( \int h(x, y) dF_{[\lambda n]+1,n}(y) \right) dF_{[\lambda n]}(x) \right. \\ &\quad \left. - \iint h(x, y) dF(x) dF(y) \right\} \\ &= [\lambda n](n - [n\lambda]) \left\{ \int \left( \int h(x, y) d(F_{[\lambda n]+1,n} - F)(y) \right) dF_{[\lambda n]}(x) \right. \\ &\quad \left. + \int \left( \int h(x, y) dF(y) \right) d(F_{[\lambda n]} - F)(x) \right\}. \end{aligned}$$

Next, we integrate by parts in order to get the “ $F_n - F$ ” terms as integrands and the deterministic terms as integrators. Regarding the first integral, we obtain

$$\begin{aligned}
 & \int h(x, y) d(F_{[\lambda n]+1, n} - F)(y) \\
 &= [h(x, y) \cdot (F_{[\lambda n]+1, n} - F)(y)]_{y=-\infty}^{\infty} - \int (F_{[\lambda n]+1, n} - F)(y) dh(x, y)(y) \\
 (39) \quad &= - \int (F_{[\lambda n]+1, n} - F)(y) dh(x, y)(y),
 \end{aligned}$$

where the boundary term vanishes, since  $h$  is bounded. Regarding the second integral, we obtain

$$\begin{aligned}
 & \int \left( \int h(x, y) dF(y) \right) d(F_{[\lambda n]} - F)(x) = [\tilde{h}(x)(F_{[\lambda n]} - F)(x)]_{x=-\infty}^{\infty} - \int (F_{[\lambda n]} - F)(x) d\tilde{h}(x) \\
 (40) \quad &= - \int (F_{[\lambda n]} - F)(x) d\tilde{h}(x),
 \end{aligned}$$

where the boundary term vanishes by boundedness of  $\tilde{h}$ . In total, we thus obtain

$$\begin{aligned}
 & \frac{1}{n d_n} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n \left( h(X_i, X_j) - \iint h(x, y) dF(x) dF(y) \right) \\
 &= \frac{[\lambda n](n - [n\lambda])}{n d_n} \left\{ \int \left( - \int (F_{[\lambda n]+1, n} - F)(y) dh(x, y)(y) \right) dF_{[\lambda n]}(x) \right. \\
 (41) \quad & \left. - \int (F_{[\lambda n]} - F)(x) d\tilde{h}(x) \right\}.
 \end{aligned}$$

We now consider both terms on the right-hand side of (41) separately, and show that they converge to the limit given in (25). Regarding the second term, we obtain

$$\begin{aligned}
 & \frac{[\lambda n](n - [n\lambda])}{n d_n} \int (F_{[\lambda n]} - F)(x) d\tilde{h}(x) - (1 - \lambda) \int J(x) Z(\lambda) d\tilde{h}(x) \\
 &= \frac{n - [n\lambda]}{n} \int (d_n^{-1} [\lambda n] (F_{[\lambda n]} - F)(x) - J(x) Z(\lambda)) d\tilde{h}(x) \\
 (42) \quad &+ \left( \frac{n - [\lambda n]}{n} - (1 - \lambda) \right) \int J(x) Z(\lambda) d\tilde{h}(x)
 \end{aligned}$$

The first summand in (42) converges to zero because of (35) and the bounded total variation of  $\tilde{h}$ . The second summand converges by (37), since  $\sup_{0 \leq \lambda \leq 1} |(n - [n\lambda])/n - (1 - \lambda)| \rightarrow 0$ .

Regarding the first term on the right-hand side of (41), we get

$$\begin{aligned}
& \frac{[\lambda n](n - [n\lambda])}{n d_n} \int \left( \int (F_{[\lambda n]+1, n} - F)(y) dh(x, y)(y) \right) dF_{[\lambda n]}(x) \\
& \quad - \lambda \int \left( \int J(y)(Z(1) - Z(\lambda)) dh(x, y)(y) \right) dF(x) \\
&= \frac{[n\lambda]}{n} \iint \left\{ d_n^{-1}(n - [n\lambda])(F_{[n\lambda]+1, n} - F)(y) - J(y)(Z(1) - Z(\lambda)) \right\} dh(x, y)(y) dF_{[n\lambda]}(x) \\
& \quad + \frac{[n\lambda]}{n} (Z(1) - Z(\lambda)) \int \left( \int J(y) dh(x, y)(y) \right) d(F_{[n\lambda]} - F)(x) \\
(43) \quad & \quad + \left( \frac{[n\lambda]}{n} - \lambda \right) (Z(1) - Z(\lambda)) \int \left( \int J(y) dh(x, y)(y) \right) dF(x)
\end{aligned}$$

All three terms on the right-hand side converge to zero. For the last term this is a consequence of  $\sup_{0 \leq \lambda \leq 1} |[n\lambda]/n - \lambda| \rightarrow 0$ , and of (38). Convergence of the first term follows from (36), together with the estimate

$$\begin{aligned}
& \left| \int \left\{ d_n^{-1}(n - [n\lambda])(F_{[n\lambda]+1, n} - F)(y) - J(y)(Z(1) - Z(\lambda)) \right\} dh(x, y)(y) \right| \\
& \leq c \sup_{\lambda, x} \left| \left\{ d_n^{-1}(n - [n\lambda])(F_{[n\lambda]+1, n} - F)(y) - J(y)(Z(1) - Z(\lambda)) \right\} \right|.
\end{aligned}$$

The second term in (43) can be written as

$$\begin{aligned}
& \frac{[n\lambda]}{n} \int \left( \int J(y) dh(x, y)(y) \right) d(F_{[n\lambda]} - F)(x) \\
&= \frac{1}{n} \sum_{i=1}^{[n\lambda]} \int J(y) dh(X_i, y)(y) - E \left[ \int J(y) dh(X_1, y)(y) \right] \\
&= \frac{1}{n} \sum_{i=1}^{[n\lambda]} (f(X_i) - Ef(X_i)),
\end{aligned}$$

where  $f(x) = \int J(t) dh(x, t)(t)$ . As  $E|f(X_i)| = \int |f(x)| dF(x) < \infty$ , we may apply the ergodic theorem to obtain convergence of the right hand side to zero, almost surely.  $\square$

## REFERENCES

- S. Bai, M.S. Taqqu (2013): Multivariate limit theorems in the context of long-range dependence, *Journal of Time Series Analysis* **34**, 717–743.
- R. T. Baillie (1996): Long memory processes and fractional integration in econometrics, *Journal of Econometrics* **73**, 5–59.
- J. Beran (2010): Long-range dependence, *Wiley Interdisciplinary Reviews: Computational Statistics* **2**, 26–35.
- E. Beutner, H. Zähle (2012): Deriving the asymptotic distribution of U- and V-statistics of dependent data using weighted empirical processes, *Bernoulli* **18**, 803–822.
- F. J. Breidt, N. Crato, P. de Lima (1998): The detection and estimation of long memory in stochastic volatility, *Journal of Econometrics* **83**, 325–348.

- P. Breuer, P. Major (1983): Central limit theorems for non-linear functionals of Gaussian fields, *Journal of Multivariate Analysis* **13**, 425–441.
- R. Caballero, S. Jewson, A. Brix (2002): Long memory in surface air temperature: detection, modeling, and application to weather derivative valuation, *Climate Research* **21**, 127–140.
- O. Cappé, E. Moulines, J.-C. Pesquet, A. Petropulu, X. Yang (2002): Long-range dependence and heavy-tail modeling for teletraffic data, *IEEE Signal Processing Magazine* **19**, 14–27.
- M. Csörgő, L. Horváth (1988): Invariance Principles for Change-point Problems. *Journal of Multivariate Analysis* **27**, 151–168.
- M. Csörgő, L. Horváth (1997): *Limit Theorems in Change Point Analysis*, J. Wiley & Sons, New York.
- H. Dehling, R. Fried, I. Garcia, M. Wendler (2013): Change-Point Detection under Dependence Based on Two-Sample U-Statistics, To appear in: *Asymptotic Methods in Stochastics - Festschrift in Honor of Miklos Csörgő's 80th Birthday* (2014+).
- H. Dehling, A. Roach, M. S. Taqu (2013a): Nonparametric Change-Point Tests for Long-Range Dependent Data. *Scandinavian Journal of Statistics* **40**, 153–173.
- H. Dehling, A. Roach, M. S. Taqu (2013b): Power of Change-Point Tests for Long-Range Dependent Data. *arXiv:1303.4917[math.ST]*.
- H. Dehling, M. S. Taqu (1989): The Empirical process of some long-range dependent sequences with an application to  $U$ -statistics, *Annals of Statistics*, **17**, 1767–1783.
- H. Dehling, M. S. Taqu (1991): Bivariate symmetric statistics of long-range dependent observations, *Journal of Statistical Planning and Inference* **28**, 153–165.
- M. Denker, M. Gordin (2013): Limit theorems for von Mises statistics of a measure preserving transformation, *Probability Theory and Related Fields* DOI 10.1007/s00440-013-0522-z.
- R. L. Dobrushin, P. Major (1979): Non-Central Limit Theorems for Non-Linear Functionals of Gaussian Fields, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **50**, 27–52.
- I.S. Gradshteyn, I.M. Ryzhik (1980): *Table of Integrals, Series and Products*, Academic Press.
- L. Horváth, P. Kokoszka (1997): The effect of long-range dependence on change-point estimators, *Journal of Statistical Planning and Inference*, **64**, 57–81.
- P.J. Huber (1981): *Robust Statistics*, John Wiley and Sons, New York.
- A. J. Lee (1990): *U-Statistics: Theory and Practice*. CRC Press, 1990.
- A. Leucht, M.H. Neumann (2013): Dependent wild bootstrap for degenerate  $U$ - and  $V$ -statistics, *Journal of Multivariate Analysis* **117**, 257–280.
- J. B. Levy, M. S. Taqu (2000): Renewal Reward Processes with Heavy-Tailed Inter-Renewal Times and Heavy-Tailed Rewards, *Bernoulli* **6**, 23–44.
- C. Lévy-Leduc, H. Boistard, E. Moulines, M. S. Taqu, V. A. Reisen (2011): Asymptotic properties of  $U$ -processes under long-range dependence, *Annals of Statistics* **39**, 1399–1426.
- C. Lévy-Leduc, M. S. Taqu (2014): Hermite ranks and  $U$ -statistics, *Metrika* **77**, 105–136.
- A. W. Lo (1991): Long-term memory in stock market prices. *Econometrica* **59**, 1279–1313.
- P. Major (1981a): *Multiple Wiener-Itô Integrals*, Lecture Notes in Mathematics 849, Springer, Berlin Heidelberg.
- P. Major (1981b): Limit theorems for non-linear functionals of Gaussian sequences, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **57**, 129–158.

- B. B. Mandelbrot, J. R. Wallis (1968): Noah, Joseph and operational hydrology, *Water Resources Research* **4**, 909–918.
- F. Móricz (1976): Moment inequalities and the strong laws of large numbers, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **35**, 299–314.
- V. Pipiras, M. S. Taqqu (2014): *Long Range Dependence and Self-Similarity*, Cambridge University Press (forthcoming).
- G. R. Shorack, J. A. Wellner (1986): *Empirical Processes with Applications to Statistics*, John Wiley & Sons, New York.
- M. S. Taqqu (1975): Weak Convergence to Fractional Brownian Motion and to the Rosenblatt Process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **31**, 287–302.
- M. S. Taqqu (1979): Convergence of Integrated Processes of Arbitrary Hermite Rank, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **50**, 53–83.
- W. Willinger, M. S. Taqqu, A. Erramilli: A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks, in: F. P. Kelly, S. Zachary, I. Ziedins (eds.) *Stochastic Networks: Theory and Applications*, Clarendon Press, Oxford, 339–366.
- W. Willinger, M. S. Taqqu, V. Teverovsky (1999): Stock market prices and long-range dependence, *Finance and Stochastics* **3**, 1–13.

FAKULTÄT FÜR MATHEMATIK, RUHR-UNIVERSITÄT BOCHUM, 44780 BOCHUM, GERMANY

*E-mail address:* herold.dehling@rub.de

*E-mail address:* aeneas.rooch@rub.de

*E-mail address:* martin.wendler@rub.de